Design and development of Novel Machine learning framework for identifying fake post

Synopsis Submitted to,

SWAMI RAMANAND TEERTH MARATHWADA UNIVERSITY

For Degree of

Doctor of Philosophy

in

Computer Science

Under the Faculty of Science and Technology

By

Mr.Khandagale Sachin Hanumant
Assistant Professor,
SangolaCollege,Sangola
M.Sc.(Computer Science), SET

Under the Guidance of

Dr.GajananD.Kurundkar
Assistant Professor,
Shri Guru BuddhiswamiMahavidyalaya,
Purna (Jn), Dist.: Parbhani-431 511 (Maharashtra) India

September 2022

Design and development of Machine Learning framework for fake information detection

Contents

1. **Introduction** 1
2. **Research Challenges** 4
3. **Objectives** 5
4. **Literature review** 6
5. **Proposed Work** 10
6. **Methodology** 10
7. **Conclusion** 15
8. **References** 16

1. Introduction

Information is one of the most essential things in our modern technological world. This world is bounded by various aspects majorly humans, an environment and multiple industries, etc. Information is the processed data that is accurately specified or organized for a particular purpose, meaning, and relevance. Information will be represented in various formats like text, image, audio or video, etc. Information sharing could be done physically or electronically.

Nowadays, Due to advanced development in hand-held devices, a large number of users are participating in Social Media Networks (SMN), which have progressively obtained information from the Internet. So, information is shared digitally with high-speed Internet. Social media has become an integral part of the Internet. Current days, social media is the basic need of every human. This digital era is growing and it plays a significant role in the day-to-day activity of human lives. Over the last few years, an increased number of users actively participated on social media every day for acquiring or sharing their thoughts, subject discussion, political views, stories, news, experiences, personal branding and views, multimedia data, etc. [1].

Peoples are just one click away from worldwide information. According to Digital Global Overview Report 2021, 4.62 billion user's active on social media platforms which are 58.4% of the total population, and the daily average time spent on media by every person is nearly 02-03 hours. These numbers represent the popularity and active involvement of people on social media. The popularity of social media is due to all-time connections with different communities, family, and friends and sharing of various kinds of information. In globally active users on popular social media platforms like 2910 million users of Facebook, 2562 million users of youtube, 2000 million users of WhatsApp, 1478 million users of Instagram, 1263 million users of WeChat, 1000 million users of TikTok, 436 million users of Twitter, etc. [2].

The above-given statistics show the use of social media and rapidly increase in the future. But a major issue of social media is the authenticity or correctness of the information. Unfortunately, most of the information is fiction (like hidden motto), not based on real facts or results. On social media platforms, most users spread information intentionally or without having proper knowledge. This is called propaganda or fake information. [3].

The spread of fake information has a potentially negative impact on individuals and society. This fake information impacts human emotions, economical factors, social factors, conspiracy, ideology, health-care options, shopping, education, opinion, perception, culture,

views, etc. Due to that reason, a human can take wrong decisions and carry misinformation. it may be harmful and risky. [3].

The spread of fake information has a serious impact on society, which may affect political election results, manipulate the stock price in global stock markets, weather crises, religious violence, bad publicity, cultural dispute, etc. In 2016, a fake information storm has affected the US election. In 2017, a large amount of spam information was generated by Socialbots during the election of the German Chancellor. Also, it happened in the French presidential election. In addition, at the time of the COVID epidemic, spreading rumors has a serious impact on health [4].

One of the major sources of fake information is a universal network of the internet which spreads incorrect and biased information easily. It is virtually impractical to prevent or control fake information from being created or disseminated. [5].

Peoples are unable to check or validate cross references of information due to lack of time. They just forward or posted information on social media blindly. This is a major issue related to information authenticity and realism.

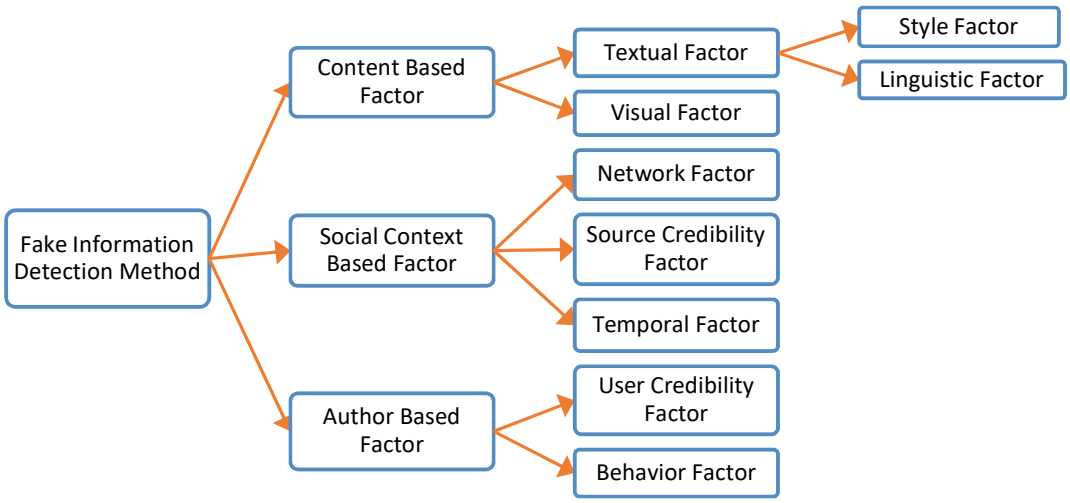
Fake information includes

1. Clickbait: stories are intentionally fabricated to gain more website visitors and increase advertising revenue for websites.
2. Propaganda: Stories are created to purposely mislead audiences, and promote a biased point of view or particular political agenda.
3. Satire/Parody: Lots of websites and social media accounts publish fake news stories for entertainment and parody.
4. Sloppy Journalism: Sometimes journalists may publish a story with unreliable information or without checking all of the facts which can mislead audiences.
5. Misleading Headings: sensationalist headlines can spread quickly on social media sites where only headlines and small snippets of the full article are displayed on audience newsfeeds.
6. Biased/Slanted News: news or stories that confirm their idea or biases. Display news based on our personalized searches.
7. Imposter Content: When genuine sources are impersonated with false, made-up sources.
8. Manipulated Content: When real information or imagery is manipulated to deceive, as with a doctored photo or video. [6]

Most information is available in textual format. To identify fake information, we need to analyze the text contents. A Text Analysis is an automatic process that uses natural language

processing (NLP) to extract valuable perceptions from unstructured text. This process is exploring and analyzing large amounts of unstructured text data and transforming it into information that machines can understand. Text Analysis is useful for identifying sentiment, contents, concepts, patterns, topics, keywords, and other attributes in the data i.e. qualitative and quantitative analysis.[7].

There are various challenges coupled with fake information. These challenges emphasize necessitate of fake information detection.[3]. Lots of work was done on fake information detection. Most researchers work on different detection methods or techniques like [8]. There is a wide research area on an effective and designed automatic framework for fake



information detection. [9].

The proposed study focuses on identifying fake information. The study suggests a hybrid model which is based on Content-based factors, Title based factors, and author-based factors. In the model, a major focus is on content, catchy titles, writing styles, the legacy or credibility of the author, and its behavior. All these factors are applied in supervised machine learning algorithms on the dataset. The dataset used for the study was downloaded from the online available dataset and is freely available.

2. Research Challenges

Nowadays,it is difficult to detect fake information. Peoples are accessing, reading, and spreading fake information until authenticate person or genuine media platform notice that they have been served false information. People are facingplentiful challenges. The recent years, the major tasks in fake information detection research will be as follows:

I. To develop an automated classification model, which is identifying fake information and unreliable authors or sourcesquickly.

II. Our research will involve training the dataset on different algorithms to determine which algorithm performs well.

III. Fake Information impact human thoughts, subject discussion, political views, Social approach, personal branding, views, etc.

IV. In this digital era, most people are depends on online content but that content may be unreliable. Identifying such unreliable content is one of the important research challenges nowadays.

3. Objectives

Machine learning is a framework that enables machines to learn from historical data without human involvement, and use that data to construct a replica for classification. Impure training data sets can't be used to create a successful classification model. These training data sets are more important than the algorithms in the Machine Learning approach in producing effective analysis results.

- To study the existing framework to detect fake information over the social network.
- To propose a metric for calculating the harmful impact of fake information.
- To design and implement the preprocessing framework for text analysis and classification.
- To study and analysis various machine learning techniques to identify the fake information.
- To identification of novel optimal feature set for solving the problem of fake information with high accuracy and better time complexity
- To propose and design a new framework to test and evaluate whether the information is fake or real.

4. Literature review

In the paper, a researcher worked on the analysis of existing fake news approaches in depth and develop a broad model with author credibility and spreading patterns of the news. In the model, a combination of various types of analysis, alerts, and detection techniques is protected users against the huge impact of fake news. For the study, the author uses Benjamin Political News Dataset, LIAR,FakeNewsNet,BuzzFeed News, etc. datasets [9].

The paper focuses on an ensemble-based deep learning model which classifies news as fake or real. The author uses the LIAR dataset. A Bi-LSTM-GRU-dense deep learning model mainly focuses on the textual attribute "Statement" and also uses the remaining attributes. NLP techniques are used for feature extraction. In the study, we achieved an accuracy of 0.898, recall of 0.916, precision of 0.913, and F-score of 0.914 respectively. The paper focused on only the statement feature that achieved the highest results. But low performance after including some other attributes like speaker profile [10].

In the article, the Author uses two steps in fake news detection. In the first step, text preprocessing is performed by Natural Language Processing (NLP) using bag-of-words, n-grams, count vectorizer, etc. techniques. In the second step, a Machine Learning classifier is used on the election dataset. Different classifier algorithms are used like Naïve Bayes, Logistic Regression, Linear SVM, Stochastic Gradient Classifier and Random Forest from the sci-kit learn library. Comparison of different classifier algorithms is done by using precision, F1 Score, and Recall parameters. The results show that SVM and logistic regression classifiers have the best performance on this dataset. The research cannot improve feature extraction or selection and fine-tuned classification [11].

A researcher emphasizes the major discussion on KWON,MediaEval,RUMDECT,RUMOUREVAL,MULTI, LIAR,PHEME, FEVER datasets for rumor detection. An effective analysis of multi-step classification and study of multiple features of rumors through the supervised and unsupervised machine learning approaches. The study uses deep learning algorithms like recursive and convolutional neural networks. The article provides a proportional analysis of ultra-modern rumor detection methods [12].

The author focused on multi-model fake information detection on textual and visual information. The text feature vectors are obtained by the BERT model and visual features are obtained by the VGG19 model. Instead of concatenating the two different types of features, A Multi-Model Compact Bilinear Pooling (MCBP) is used to combine both visual and text feature vectors. The dataset was downloaded from Twitter and Weibo for validating the MCBP method. Study shows that the single MCBP method is better than the dual MCBP method. Different evaluation majors like accuracy, precision, recall, and F1 scores are used [4].

In the article, the author uses a combination of word vector representations with deep learning models. It develops Glove Feed Forward Network (FNN) and Long Short-Term Memory (LSTM) for fake news detection. These models are paired with a live data mining component, which extracts supplemental data from the headline and body of news articles. Other features such as domain names, author information, etc. are added to the original article before the word embedding stage. George McIntire's Fake News Dataset was used for testing. According to the classification of results, a significant enhancement in precision, recall, accuracy, and F1 score with the combination of LSTM and word2vec representation [3].

A researcher dedicated a blended machine learning ensemble model that was built from logistic regression, support vector machines, linear discriminant analysis, stochastic gradient descent, and ridge regression. The LIAR and ISOT datasets are used for data analysis. These datasets are publicly available to predict whether the news is real or fake. Firstly, the dataset was preprocessed using NLP and extraction of features from news texts. This cleaned dataset is passed to the blending ensemble method to improve accuracy. In the paper, the performance analysis of traditional machine learning and blending ensemble models is based on ROC AUC, AUC, precision, recall, and accuracy. The result shows ROC AUC, AUC, recall, and accuracy had the best results. The limitations of the blended machine learning ensemble model are often inconsistent, which leads to errors or anomalies in the predictions [13].

In the paper, a deep learning technique is used, along with pre-trained word embedding, namely Word2Vec, GloVe, and fastText. It has trained on four different datasets. The author uses deep learning algorithms like CNN, Bidirectional LSTM, and ResNet. To reduce data imbalances between classes, each data set undergoes data enhancement using the back-

translation method. The paper used the fastText Word embedding model with the LSTM bidirectional classification model and returned results of 99.24% accuracy, 99.19% Precision, 99.26% recall, and 99.23% F1 score. The result shows that the Bidirectional LSTM architecture outperformed CNN and ResNet. According to the results, data enhancement is effective in improving the consistency of model performance. LIAR dataset is used [14].

In the paper, the author studied different content-based and social context-based Fake news detection methods. The content-based approach assumes that the article or post contains textual or visual content or both. A content-based approach to identifying fake news in its early stages. The social context-based approach takes into account publisher structure and publisher reputation.But social context-based approaches are unable to identify fake news in the early stages because they lack the propagation details [8].

The author built a new model for fake news detection using Python. They use one of the most popular machine learning modules scikit learn. The module can provide different machine learning features such as tokenization, feature extraction, Count Vectorizer, Tiff Vectorizer, etc. XGBOOST algorithm shows the highest accuracy with more than 75%, whereas Support Vector Machine and Random Forest algorithms show around 73% accuracy [15].

In the article, the ensemble techniques used various linguistic feature sets to classify news articles from multiple domains as true or fake. The author uses various ensemble techniques such as bagging, boosting, and voting classifier and evaluates the performance over the multiple datasets. For model generation, linguistic features need some textual properties to be transformed into a numerical form. To maximize the accuracy of a dataset it needs to maintain an ideal balance between variance and bias, for that purpose various hyperparameters were used. On dataset1 (ISOT Fake News Dataset), ensemble learners achieve an average accuracy of 97.67% and 95.25% for individual learners. A random forest algorithm and Perez-LSVM achieve 99% accuracy on dataset1. The accuracy of linear SVM, multilayer perceptron, bagging classifiers, and boosting classifiers was 98%. On dataset2, linear SVM, random forest, and Perez-LSVM performed poorly. Wang-CNN and Wang-Bi-LSTM performed worse than all other algorithms [16].

Researchers projected a novel hybrid deep learning convolutional and recurrent neural network models used for fake news classification. For experimental purposes, the author uses

ISOT and FA-KES datasets. It has provided better detection results than non-hybrid baseline methods. These models perform well on the above dataset, but other dataset works poorly. Overall, using artificial neural networks to detect fake news is promising. In addition, CNNs and RNNs are more complicated neural network architectures that will be examined in future research [17].

The article uses artificial intelligence, natural language processing, and machine learning concepts to perform binary classification on various news articles. With using of grid search parameter optimization of the Logistic Regression model the accuracy was increased from 65% to 75%. The result of the study shows that, if a user passes a particular news article or its headline, there are 75% chance to classify true or fake nature [18].

The paper focused on a deep learning framework like neural networks and long short-term memory architecture for the detection of fake news. The author has extracted feature with help of the GloVe approach for word embeddings and tokenization. The long short-term memory model is primarily focused on data classification for sequential prediction. The model achieved 99.88% accuracy. The goal was to expand the current work to create an automated system [19].

5. Proposed Work

The proposed study focuses on identifying fake information. The study suggests a hybrid model which is based on Content-based factors, Title based factors, and author-based factors. In the model, a major focus is on content, catchy titles, writing styles, the legacy or credibility of the author, and its behavior. All these factors are applied in supervised machine learning algorithms on the dataset. The dataset used for the study was downloaded from the online available dataset and is freely available.

6. Methodology

Holistic development in the social media network field enables sharing their personal experiences, feelings, opinions, sentiments, thoughts, and idea's on different topics or current issues. This information can impact a user's decisions such as views, economic, social, political, religious, educational, psychological, etc. Thus, most users habitually depend on public opinion to reform their decisions. But shared experiences, opinions, sentiments, thoughts, and ideas posted on news, blog, post, and review may be fake. Fake content is used for advertising, promotion, disseminating information, and for financial gain.

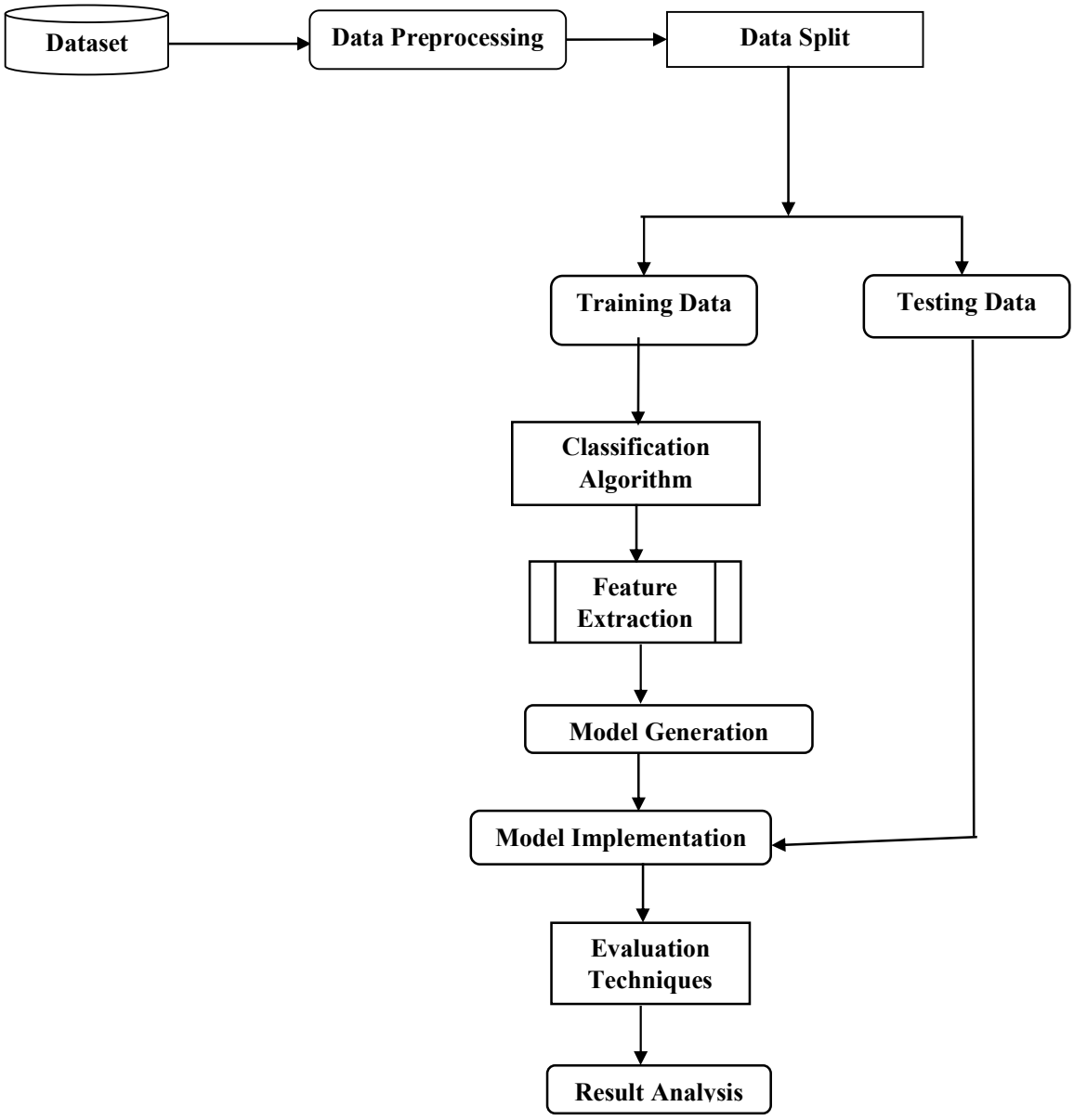
To identify fake information, the First thing is always to check the content of the information whether it's reliable or trusted. Second, access only reputed knowledgeable author's information. Third, information is accessed from only recognized sources and authorized websites or platforms.

The proposed study focuses on identifying fake information. The study suggests a hybrid model which is based on Content-based factors, Title based factors, and author-based factors. In the model, a major focus is on content, catchy titles, writing styles, the legacy or credibility of the author, and its behavior. All these factors are applied in supervised machine learning algorithms on the dataset. The dataset used for the study is downloaded from the online available website and is freely available.

I. Dataset:

In this research work, datasets play a major role in fake information detection. Dataset is a collection of records or information which is from different fields. Dataset is in unstructured

textual format and downloaded from the online available free sources. It may contain collected news or social media post, reviews, etc. The dataset contains a title or headline, a body of message or text, author information like author id, name, profile, writing style or pattern, and also a source of information.



II. Data preprocessing:

Data in the real world is often quite messy, inconsistent, noisy, or incomplete information and missing values. Data Preprocessing is needed because unstructured data convert into a human understandable and readable format. There are many data preprocessing methods such as clean, balance, replace, import, partition, scale, enhancing, and unbiased data. This helps to formulate and construct primitive data and is also suitable for building training machine

learning models.Data preprocessing has some major steps before constructing a model with these features.

- **Data quality assessment:** Based on high-quality data without noise
- **Data cleaning:** a process of correcting, fixing, or removing wrong or irrelevant data from a datasetsuch as punctuation removal,stop word removal, andstripping.
- **Data transformation:**method of converting data into proper format.
- **Data reduction:** reduce the size of data and drop the data analysis cost [20].

In Text analysis, preprocessing converts text into a clean, consistent format that can be fed into a machine learning model. Data preprocessing contains

1. Convert text into lowercase letters.
2. Remove punctuation symbols, digits, and special characters from the text.
3. Remove unnecessary spaces.
4. Remove stop words from the text. e.g. for, in, if, not in, and, we, are, you, etc.
5. Removing the prefix or suffix of a word (stemming). e.g. –ing,-ed, -size, -de, -s, etc.
6. To convert a word to its original form(Lemmatization)
7. Word tokenization divides each sentence into individual words and paragraphs into sentences.
8. Word Embedding- Feature extraction for text classifications,Compute similar words, and Document clustering/grouping

Output accuracy depends on inputted data. If the input is a fault or incomplete then the data is not correctly clean. It may incorporate faulty data.

III. Data Split:

Data is usually split in an 80:20 ratio. That is, 80% is training data and the rest is testing data. A model should learn from a training set and optimize its parameters based on the observations made. Test data are the set of data that are compared to previous test data in the final model. the model predicts a certain value and compares it to the actual result to calculate the accuracy. A testing set evaluates the final mode and algorithm. Training data and testing data both are different.

IV. Feature Extraction:

Datasets can be a large number of variables that take a lot of time to select or combine into features. Feature extraction is a dimensionality reduction process that reduces the initial set of

raw data into manageable groups for processing. After effective reduction, an accurate and completely describing the original data set. It converts text into a vector of features.

A Word vectorization technique is categorized like

- **Bag of Words (BoW):**

  Text and documents can be analysed based on word count.

- **Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer:**

  Less frequently occurring words are given more weight.

  - **Term Frequency (TF)-**Provides a count of how many times a word appears in the document.

$$TF = \frac{Frequency\ of\ word\ in\ a\ document}{Total\ number\ of\ words\ in\ that\ document}$$

  - **Inverse Document Frequency-** Analysis of whether a term occurs rarely or frequently in documents across the corpus

$$DF = \frac{Documents\ containing\ word\ W}{Total\ number\ of\ documents} \qquad IDF = \log\left(\frac{Total\ number\ of\ documents}{Documents\ containing\ word\ W}\right)$$

  The Result of TF-IDF = TF * IDF

- **<u>CountVectorizer</u>**

  Used to convert a particular text to a vector based on the frequency (count) of each word in the text.

V. Classification Algorithm:

There are different types of classification algorithms, each with its unique function and application. The program you use for a particular task depends on the purpose of the task and the data you want to retrieve. The objective of classification is to accurately predict the target class for each record in the data.

We have various classification methods in machine learning like Logistic regression, Linear regression, Decision trees, Random forest, Naive Bayes, Support Vector Machines, K-nearest neighbors, and deep learning are used. The best classification algorithm and feature extraction method are used to indicate whether the information is fake or real.

VI. Working of Classification:

Classification is a process of finding a model that involves classifying the new observations based on observed patterns from the previous data. In classification, the dataset is divided into training and testing data. Classification constructs the model by using training data set. The algorithm predicts the value or class label of unseen data based on the model.

The classification technique works in two steps:-

a. Model Generation and
b. Model Implementation

a. Model Generation:

In the classification dataset, each row is assigned a predefined class label. Different machine learning algorithms are used to build a classifier by analyzing a training set. A training dataset contains the different attributes and associated class labels. The algorithm uses the training data set to create a model or classifier.

There are many individual measurable properties and features associated with a phenomenon that can be observed and measured. The main objective of the classification algorithm is to identify the class or group of a particular data set, and these algorithms are generally used to forecast the output. On a given testing data, the accuracy of a classifier is calculated by the percentage of records that are correctly classified.

b. Model Implementation:

A constructed model is utilized to classify unseen data. After generating the classification model, we can input data that we need to predict, adopt the same flow, and feed it into our model. Upon completion, we will receive our predicted class.

H/W and S/W Requirements:-

This model is implemented in python and anaconda. For implementation purposes, different libraries of python are used. The libraries which are used for implementation purposes are nltk, re, textblog, NumPy, etc.

VII. Evaluation Techniques

The evaluation technique helps you to assess your model's performance, monitor, and control. We aim to create a model which gives high accuracy on out-of-sample data. Evaluation metrics are tied to machine learning tasks. Using different metrics for the tasks of classification describes the performance of a model on a set of test data. Correct and incorrect predictions are grouped by class.

a. Accuracy: the percentage of correct predictions for the test data. the ratio between the number of correct predictions to the total number of predictions.

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

i.e. Accuracy= number of correct predictions / Total number of predictions

b. Precision: To identify the correctness of classification. The ratio between the number of *Positive* samples correctly classified to the total number of samples classified as *Positive* (either correctly or incorrectly). Higher precision means better the ability of the model to correctly classify the positive class.

$$Precision = \frac{True_{positive}}{True_{positive} + False_{positive}}$$

i.e. Precision = Number of correct positive results / Total number of positive results

c. Recall: Recall tells us the number of positive cases correctly identified out of the total number of positive cases.A high recall value indicates that a lot of fake cases are identified out of the total number of fakes.

$$Recall = \frac{True_{positive}}{True_{positive} + False_{negative}}$$

Recall = Number of correct positives/number of all positives

***d.* F1- Score:** weighted average of precision and recall. Precision is used to save the company's costs. The recall is used to ensure the machinery is stable and not a threat to human lives[22].It is a measure of a test's accuracy

$$F1 - score = 2 * \frac{precision * recall}{precision + recall}$$

In machine learning, these metrics are commonly used to evaluate classifier performance from different angles. Accuracy measures the similarity between fake information predictions and actual fake information.

VIII. Result Analysis

Standard metrics measure the overall performance of each algorithm in the classification process. A matrix illustrating the results of each classifier on the test data is shown by the number of correct and incorrect predictions for each class. Each model's performance was is evaluated in detail to provide a comprehensive result analysis. This study aims to compare all algorithms in terms of their overall performance.

7. Conclusion

The rapidly increasing use of social media platforms for circulating information, post, reviews, news, etc. It pushes research in fake information detection to identify contents,

authors, and sources of information. Information is quickly spread on online platforms without checking its authenticity.

So this Research provides a model which is used to identify the authenticity of the information. For detecting fake information this model will be used text analysis technique. In the proposed study, contextual, title, and author features will be used to detect fake information. In the study, different machine learning algorithms will be used to identify fake information effectively and efficiently. This study focuses on the best machine learning algorithm for fake information detection based on different evaluation measures.

8. References

[1] [https://www.theglobalstatistics.com/india-social-media-statistics/, accessed on 4/8/2022].

[2] https://datareportal.com/reports/digital-2022-global-overview-report

[3] Deepak S, BhadrachalamChitturi, "Deep neural approach to Fake-News identification", International Conference on Computational Intelligence and Data Science (ICCIDS 2019), 2020, Elsevier B.V, Procedia Computer Science, 167pp, 2236–2243

[4] Ning Xiang, "Deep Learning-Based Fake Information Detection and Influence Evaluation", 2022, Article ID 8514430, Hindawi Computational Intelligence and Neuroscience, pp 8

[5] Arvin Hansrajh, Timothy T. Adeliyi, and Jeanette Wing, "Detection of Online Fake News Using Blending Ensemble Learning", 2021, Article ID 3434458, Hindawi Scientific Programming, pp 10

[6] https://www.webwise.ie/teachers/what-is-fake-news/, accessed on 4/8/2022

[7] https://monkeylearn.com/text-mining/, accessed on 4/8/2022

[8] Sakshini Hangloo, Bhavna Arora, "FAKE NEWS DETECTION TOOLS AND METHODS – A REVIEW", April - June 2021, International Journal of Advance and Innovative Research, Volume 8, Issue 2 (IX), ISSN 2394 - 7780

[9] Xichen Zhang, Ali A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion", 8 March 2019, Information Processing and Management, Elsevier

[10] NidaAslam, IrfanUllah Khan, Farah Salem Alotaibi, Lama AbdulazizAldaej, and Asma Khaled Aldubaikil, "Fake Detect: A Deep Learning Ensemble Model for Fake News Detection", 2021

[11] Vasu Agarwal, H.Parveen Sultana, Srijan Malhotra, Amitrajit Sarkar, “Analysis of Classifiers for Fake News Detection”, 2019, INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, Procedia Computer Science, 165, 377–383, Elsevier Ltd

[12] Ajeet Ram Pathak, Aditee Mahajan, Keshav Singh, AishwaryaPatil, Anusha Nair, “Analysis of Techniques for Rumor Detection in Social Media”, International Conference on Computational Intelligence and Data Science, 2019, Procedia Computer Science 1877-0509, Elsevier B.V

[13] Arvin Hansrajh, Timothy T. Adeliyi, and Jeanette Wing, “Detection of Online Fake News Using Blending Ensemble Learning”, 2021, Article ID 3434458, Hindawi Scientific Programming, pp 10

[14] I. KadekSastrawan, I.P.A. Bayupati, Dewa Made Sri Arsa, “Detection of fake news using deep learning CNN–RNN based methods”, Oct 2021,Korean Institute of Communications and Information Sciences,Elsevier B.V.

[15] Z Khanam, B N Alwasel, H Sirafi, M Rashid, “Fake News Detection Using Machine Learning Approaches”,2021,IOP Conf. Series: Materials Science and Engineering,1099 (2021) 012040,IOP Publishing

[16] Iftikhar Ahmad, Muhammad Yousaf, SuhailYousaf, Muhammad Ovais Ahmad “Fake News Detection Using Machine Learning Ensemble Methods”,17 October 2020,Hindawi Complexity Volume 2020, Article ID 8885861, 11 pages

[17] Jamal Abdul Nasir, Osama Subhani Khan, IraklisVarlamis, “Fake news detection: A hybrid CNN-RNN based deep learning approach”,29 December 2020,2667-0968,International Journal of Information Management Data Insights 1 (2021),Elsevier Ltd.

[18] Uma Sharma, Sidarth Saran, Shankar M. Patil, “Fake News Detection Using Machine Learning Algorithms”,International Journal of Creative Research Thoughts (IJCRT), 6 June 2020, Volume 8, ISSN: 2320-2882, 1394-1402 pp

[19] Tavishee Chauhan, HemantPalivela, “Optimization and improvement of fake news detection using deep learning approaches for societal benefit”, 2021,International Journal of Information Management Data Insights 1 100051,Elsevier Ltd

[20] https://serokell.io/blog/data-preprocessing, accessed on 23 Aug 2022

[21] Dipak R. Kawade, Kavita S. Oza, “News Classification: A Data Mining Approach”, Indian Journal of Science and Technology, Vol 9(46), ISSN (Online): 0974-5645, December 2016

Signature **Signature**

(Research Scholar) **(Research Guide)**
Mr. Sachin H. KhandagaleDr.GajananD.Kurundkar
Assistant Professor,
Shri Guru BuddhiswamiMahavidyalaya,
Purna (Jn), Dist.: Parbhani-431 511